SPECULAR HIGHLIGHT ENHANCEMENT FROM VIDEO SEQUENCES

Véronique Prinet *Michael Werman* *Dani Lischinski*

The Hebrew University of Jerusalem, Israel

ABSTRACT

We propose a novel method for detecting and enhancing specular highlights from a video sequence, thereby obtaining a better visual perception of the specularity. To this end, we first generate a specularity map, defined over the space-time domain, by leveraging information provided by the temporal acquisition. We then amplify the highlights in each video frame to create the visual sensation of high dynamic range data. Results are illustrated on several videos taken under different acquisition scenarios.

Index Terms— Image and video processing and enhancement, Feature extraction, Image and video representation and rendering.

1. INTRODUCTION

Specular highlights in images or videos result from the mirror-like reflection of incident light on a surface. Because highlights often appear as saturated areas (or areas where the object intrinsic spectral reflectance properties disappear under the specularity), they can cause vision algorithms for segmentation, shape from shading, or motion estimation, to produce erroneous results. For this reason, highlight removal approaches have been introduced since the early years of image processing and computer vision [1]. More recently, it has been shown that the presence of highlights or direct light sources can be exploited advantageously to enhance the visual quality of the images/videos, or to produce High Dynamic Range data from single-exposure Low Dynamic Range inputs [2, 3, 4, 5].

The contribution of this paper is twofold: (i) we propose to estimate the *specularity amplitude* ('degree' of highlight) at each pixel of the video, without hard thresholding ; this differs from previous approaches that mostly detect the highlights/light sources as the brightest points/regions in the image/video [6, 4]; (ii) we magnify the resulting specularity map using a simple sigmoid function. Though this choice is a priori heuristic, it enables the enhancement of local contrast around strong highlights, conveying a better visual perception of the specularity.

2. RELATED WORK

Specular highlight detection from multiple images.

Multi-illuminants (multi-polarisation, multi-flash, multi-frequency), static camera. Nayar et al. [7] proposed a technique for separating the specular and the diffuse component of reflection from images, using color and polarisation information simultaneously. In [8], Feris et al. use a multi-flash camera to take multiple pictures of the scene, each with a differently positioned light source (flash).

Multi-views (stereo-like), single illuminant. Lee et al. [1] developed an algorithm based on the observation that the diffuse component of colour image irradiance does not change depending on viewing directions, while the specular reflection does change. The spectral differencing detects specularities by pixelwise colour differences. Extending this work, Lin et al. [9] proposed a form of 'color histogram differencing' that utilizes the epipolar constraint from multi-baseline stereo images. Recently, Yang [10], introduced a framework for simultaneous estimation of point correspondence and specular reflection from pairs of stereo images. A general survey on specularity removal techniques can be found in [11].

Inverse Tone Mapping Operators.

Rempel et al. [2] introduced a LDR to HDR algorithm for real-time processing of images and video streams. The approach is based on a brightness enhancement function, that enables to expand the range of bright/saturated areas (detected by thresholding), locally and in a non-linear way. Similarly, Banterle et al. [4] propose combining a global inverse tone mapping algorithm with a local expansion map. It makes it possible to boost bright regions in the image without affecting darker regions. The processing can be performed in spatial and temporal domains to avoid possible temporal flickering in videos. The system proposed by Didyk et al. [12] consists of classifying clipped (saturated) regions as lights, reflection or diffuse surfaces, using a semi-automatic classifier; it then enhances each class of objects differently. Application is to video sequences.

The assumptions underlying to most algorithms cited above pose limitations on the applicable domains of the objects and illumination. One of the main motivations of the present research is to develop a model applicable to more general conditions, leveraging temporal information.

3. IMAGE FORMATION MODEL AND BASIC ASSUMPTIONS

3.1. Shafer's reflection model

The dichromatic model for dielectric objects (such as plastic, acrylics, hairs, etc.) models the light reflected from an object as a linear combination of diffuse and specular reflectance as follows [13]:

$$\mathbf{J}(\mathbf{p}) = \mathbf{D}(\mathbf{p}) + \mathbf{S}(\mathbf{p}) \qquad \forall \mathbf{p} \in \Omega. \tag{1}$$

$\mathbf{J} = \{J_c\}$, $\mathbf{D} = \{D_c\}$, $\mathbf{S} = \{S_c\} \in \mathbb{R}^+, \forall c \in \{r, g, b\}$, are respectively the observed irradiance function, the diffuse reflectance function and the specular reflectance function; $\mathbf{p} = (x, y) \subset \Omega$ is a pixel in the image.

- The diffuse term (or body component) follows Lambert's law and characterizes the object's spectral properties. It is usually modelled as a product between the object albedo (or *intrinsic image*), $\mathbf{R}$, and the incident illuminant $\mathbf{L} = \{L_c\}$: $D_c(\mathbf{p}) = m_d(\mathbf{p})R_c(\mathbf{p}) L_c(\mathbf{p}), \forall c \in \{r, g, b\}$. The weighting scalar $m_d(\mathbf{p})$ is function of the incident light direction and the surface normal at $\mathbf{p}$ [14].

- The specular term (or interface reflectance) is associated with the directional reflection of the incident light beam hitting the object's surface. It is commonly assumed that its spectral distribution is identical to the spectral distribution of the incident light [13], hence can be expressed: $S_c(\mathbf{p}, t) = m_s(\mathbf{p}, t) L_c, \forall c \in \{r, g, b\}$, where $m_s(\mathbf{p})$ varies as an exponential law with the angle formed between the surface normal and the view angle [15].

If the global incident light source is white and uniform in time and space, $\mathbf{L}(\mathbf{p}, t) = \mathbf{L} = (1, 1, 1)$, Shafer's model, for a time series, becomes:

$$\mathbf{J}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p}, t) + m(\mathbf{p}, t) \, \mathbf{I}_{3 \times 1} \qquad \forall \mathbf{p} \in \Omega(t), \tag{2}$$

$\forall t \in \mathbf{V}$, where $\mathbf{V} = \{t_1, .. t_n\}$ denotes the time sequence, and $\mathbf{I}_{3 \times 1}$ is a unit vector of length three; $m(\mathbf{p}, t) = m_s(\mathbf{p}, t)$ is the *amplitude of the specularity* at point $(\mathbf{p}, t)$. Equation (2) is the image formation model under a white illuminant.

3.2. Assumptions

White global illuminant. For input images or videos that exhibit a certain color cast, white balance correction can be applied using existing techniques [16]. Most diffuse-specular separation algorithms from single images based on Shafer's model make a similar assumption [17].

Uniform body refectance over time. For a fixed physical point $\mathcal{P}$ of the 3D world projected to $\mathbf{p}$ in frame $\Omega(t)$, and to $\mathbf{p} + \mathbf{\Delta p}$ in frame $\Omega(t + \Delta t)$, where Δt is a small increment of time, its diffuse component $\mathbf{D}$ does not change under constant illumination:

$$\mathbf{D}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p} + \mathbf{\Delta p}, t + \Delta t). \tag{3}$$

This assumption was used by Lee et al. [1] and related work. It is likely to hold in any situations where the camera/viewer is in motion, while the light source and objects in the scene are static (since $\mathbf{D}(\mathbf{p})$ is camera position independent). We will show that, experimentally, it is also a reasonable assumption in scenes where light sources are moving.

Known displacement $\mathbf{\Delta p}$. We compute the motion field from frame to frame, $\mathbf{\Delta p}(\mathbf{p}, t)$, using optical flow technique [18] ; we use this flow as a 'true displacement'. Frame to frame motion is assumed to be reasonably small. Alternatively, in [9], the authors estimate point correspondance from stereo.

4. SPECULARITY MAP

The first step is to estimate a specularity map, defined over the entire video domain.

4.1. Formulation

From equations (2) and (3), the total derivative of the irradiance, for each channel c, may be expressed as:

$$J_c(\mathbf{p} + \mathbf{\Delta p}, t + \Delta t) - J_c(\mathbf{p}, t) \tag{4}$$
$$= m(\mathbf{p} + \mathbf{\Delta p}, t + \Delta t) - m(\mathbf{p}, t).$$

The diffuse term has cancelled out due to property (3). This equation expresses that changes in the observed irradiance of a given object tracked over time, originate from the variation in amplitude of the specular component reflected by this object. Note that m is achromatic, i.e. independent of channels c. We will show how to estimate the scalar field $m : \Omega \times \mathbf{V} \to [0, 1], m \subset C^1$.

4.2. Integration

We first derive a solution when objects in the scene and camera are static ($\mathbf{\Delta p} = 0$), while the light source(s) change [19, 7, 8]. We then generalise to $\mathbf{\Delta p} \neq 0$ [9, 10].

Case 1 (static camera/object): $\mathbf{\Delta p} = 0$
We can treat $\mathbf{p} = const$ as a parameter of $\mathbf{J}()$ and $m()$, now functions of a single variable t. If we approximate local derivatives by finite differences, then equation (4) writes:

$$\frac{dm(t)}{dt} = \frac{dJ_c(t)}{dt} \qquad \forall t \in \mathbf{V}, \tag{5}$$

with $\Delta t = 1$. The analytical solution of this first order ODE is trivially given by : $m(\mathbf{p}, t) = \mathbf{J}(\mathbf{p}, t) + \mathbf{C}(\mathbf{p})$. From initial

conditions at $t = t_0$, the constant term writes: $\mathbf{C}(\mathbf{p}) = m(\mathbf{p}, t_0) - \mathbf{J}(\mathbf{p}, t_0) = -\mathbf{D}(\mathbf{p})$. The time-independent term $\mathbf{D}(\mathbf{p})$ can be approximated as the image with smallest specularity amplitude over the time sequence: because $\min_t \mathbf{J}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p}) + \min_t m(\mathbf{p}, t)$ with $0 \leq m(\mathbf{p}, t) \leq 1$, we can define $\tilde{\mathbf{D}}(\mathbf{p}) \triangleq \min_t \mathbf{J}(\mathbf{p}, t)$ (which is not necessarily the 'true' diffuse image). Hence :

$$m(\mathbf{p}, t) = J_c(\mathbf{p}, t) - \min_t J_c(\mathbf{p}, t) \quad \forall c \in \{r, g, b\}. \quad (6)$$

Case 2 (objects/camera motion) : $\Delta \mathbf{p} = \Delta \mathbf{p}(\mathbf{p}, t)$
In a general situation, we can reduce the problem to a 1D case. Applying change of variable $\mathbf{X} = (\mathbf{p}, t)$, and again, approximating finite difference by derivative, equation (4) becomes:

$$\frac{dm(\mathbf{X})}{d\mathbf{X}} \Delta \mathbf{X} = \frac{dJ_c(\mathbf{X})}{d\mathbf{X}} \Delta \mathbf{X}. \quad (7)$$

Here $\frac{df}{d\mathbf{X}}$ is the directional derivative of f along $\Delta \mathbf{X} = (\Delta \mathbf{p}, \Delta t)$ for $\Delta \mathbf{X} \to 0$.

The integration, from $\mathbf{X}_0 = (\mathbf{p}(t_0), t_0)$ to $\mathbf{X}$, on both sides of equation (7), again leads to:

$$\begin{aligned} m(\mathbf{X}) &= m(\mathbf{X} - \Delta \mathbf{X}) + \Delta J_c(\mathbf{X} - \Delta \mathbf{X}) \\ &= J_c(\mathbf{X}) - D_{0,c} \quad \forall c \in \{r, g, b\}, \end{aligned} \quad (8)$$

with $D_{0,c} = J_c(\mathbf{X}_0) - m(\mathbf{X}_0)$.

To estimate $m_0 = m(\mathbf{X}_0)$, we proceed in two steps: (i) forward integration is first performed from t_0 to t_n by setting $m_0 = 0$; at each time frame t, the estimated specularity is clipped so that $0 \leq m(\mathbf{p}, t) \leq 1$; this process ensures that $\min_t m(\mathbf{p}(t), t) \geq 0$; (ii) backward integration, from t_n to t_0, is then applied, using for initialisation $m(\mathbf{p}, t_n), \forall \mathbf{p} \in \Omega(t_n)$. The values $m(\mathbf{p}, t_n)$ are likely to be a correct estimation of the true specularity amplitude at t_n, enabling the computation of the correct offset $D_{n,c}$; hence, integration from t_n to t_0 will furnish the final specularity map.

5. AMPLIFICATION

We amplify the highlights in each video frame to create the visual sensation of high dynamic range data.

Video amplification. We blend an enhanced specular map with the original input video, frame by frame. Our image composition rule applies as follows:

$$\tilde{\mathbf{J}}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p}, t) + m(\mathbf{p}, t)\ (k + \alpha\ g_{\beta,\theta}(m_{\mathbf{p},t}))\ \mathbf{I}_{3\times 1}$$
$$\text{with} \quad g_{\beta,\theta}(x) = 1/(1 + \exp^{-\beta(x-\theta)}),$$

where $\tilde{\mathbf{J}}$ is the output enhanced video. Parameter α controls the magnitude of amplification, β is related to the relative rate of growth of $g()$, θ is a threshold above which the signal x is amplified, under which it is attenuated; and $k \in \{0, 1\}$.

Spatial and Temporal filtering. For sequences with camera or object motion (case 2), we apply bilateral filtering to the specular map computed from equation (8). The filtering is applied, in space domain, at each time step t, using the input frame as a guiding image. This process enables the removal/reduction of spurious detections due to possible noise or incorrect displacement estimation from optical flow. In 'static' videos (case 1), we instead remove specularity components of high temporal frequency, by applying a low pass filter in the Fourier space. In both cases, the filtering is applied prior to highlight amplification.

6. EXPERIMENTAL RESULTS

We applied our approach on several data sequences acquired under three different conditions: static scene and dynamic light source (case 1), camera motion and static light source (case 2), camera motion and object motion with uncontrolled illumination conditions. The video sequences were captured with a high definition video camera (Panasonic HD-TM 700), at 60 frames per second and 1920×1080 pixels resolution.

The input video frames are first linearised by applying gamma correction, with $\gamma = 2.2$, then white balance is performed. For display, the enhanced videos after amplification $\tilde{\mathbf{J}}$, are clipped to one.

We illustrate here three results. Illustrations are best viewed in full resolution (zoom). In Figure 1, top left and right, are respectively one representative frame of the input sequence (total of 60 frames), and the reconstructed frame with enhanced highlights (with $\alpha = 4, \beta = 10, \theta = 0.1$). The time-cumulative specularity map is shown on bottom row. Note in particular the enhanced reflection of the light in the window. Figure 2 illustrates, on left side, a frame of an input sequence (60 frames), taken under fixed point light source, while the camera is moving (translation motion). The coffee bag and tea box have localised highlights which appear more shiny on the enhanced frames (right image ; $\alpha = 8, \beta = 30, \theta = 0.1$). Figure 3 illustrates five temporal frames extracted from a video sequence (total 30 frames) at regular interval, where people and camera are in motion. The bottom row show the result of specular highlight amplification ($\alpha = 5, \beta = 20, \theta = 0.2$) used in conjunction with content-aware video enhancement technique [20]. Highlight are particularly noticeable on the ground.

7. CONCLUSION

In this work, we proposed a novel method to estimate and enhance specular highlights from temporal sequences. We illustrated results on several video sequences.

Fig. 1. **Top row left**: One frame of a input video sequence (total length of 60 frames). **Top row right**: Enhanced video frame. **Bottom row**: Time cumulative specular map.

Fig. 2. **Left**: One frame of a input video sequence (total length of 60 frames). **Right**: Enhanced video frame.

Fig. 3. Time frames extracted from a video (1,5,10,15,20). **Top row**: Input data. **Bottom row**: Specular highlight enhanced.

8. REFERENCES

[1] A. W. Lee and R. Bajcsy, "Detection of specularity using colour and multiple views," *Image and Vision Computing*, 1992.

[2] A. Rempel, M. Trentacoste, H Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2hdr: On-the-fly reverse tone mapping of legacy video and photograph," in *ACM Siggraph*, 2007.

[3] L. Wang, L.Y. Wei, K. Zhou, B.N. Guo, and H.Y. Shum, "High dynamic range image hallucination," in *Eurographic Symposium on Rendering*, 2007.

[4] F. Banterle, P. Ledda, and A. Chalmers, "Expanding low dynamic range videos for hdr applications," in *SCCG*, 2008.

[5] A. Yoshida, R. Mantiuk, K. Myszkowski, and H.P. Seidel, "Analysis of reproducing real-world appearance on displays of varying range," in *Eurographics*, 2006.

[6] L. Meylan, S. Daly, and S. Susstrunk, "Tone mapping for hdr range display," in *Electronic Imaging*, 2007.

[7] S.K. Nayar, X.S. Fang, and R. Boult, "Separation of reflection components using color and polarization," *Int. J. Comp. Vision*, 1997.

[8] R. Feris, R. Raskar, K. Tan, and M. Turk, "Specular reflection reduction with multi-flash imaging," Tech. Rep., Mitsubishi Electric Research Lab., 2004.

[9] S. Lin, Y.Z. Li, S.B. Kang, X. Tong, and H.Y. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *ECCV*, 2002.

[10] Q.X. Yang, S.N. Wang, and R.G. Yang, "A uniform framework for estimating chromaticity, correspondence and specular reflection," *IEEE TIP*, 2011.

[11] A. Artusi, F. Banterle, and D. Chetverikov, "A survey of specularity removal methods," *Computer Graphics Forum*, 2011.

[12] P. Didyk, R. Mantiuk, M. Hein, and H.P. Seidel, "Enhancement of bright video features for hdr displays," in *Eurographic Symposium on Rendering*, 2008.

[13] S. Shafer, "Using color to separate reflection components," *Color Research Applications*, 1985.

[14] Y. Weiss, "Deriving intrinsic images from image sequences," in *ICCV*, 2001.

[15] K.E. Torrance and E.M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *J. Opt. Soc. Am*, 1967.

[16] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: survey and experiments," *IEEE TIP*, 2011.

[17] R. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE PAMI*, 2005.

[18] C. Liu, W.T. Freeman, E.H. Adelson, and Y. Weiss, "Human-assisted motion annotation," in *CVPR*, 2008.

[19] S.K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar, "Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination," *ACM TOG*, 2006.

[20] L. Kaufman, D. Lischinski, and M. Werman, "Content-aware automatic photo enhancement," *Computer Graphics Forum*, 2012.